

PERCEPTUAL CENTRE CORRELATES IN EVOKED POTENTIALS

C. Soraghan*, T. Ward*, R. Villing* and J. Timoney**

* Dept. of Electronic Engineering, NUI Maynooth, Maynooth, Ireland

** Dept. of Computer Science, NUI Maynooth, Maynooth, Ireland

christopher.j.soraghan@nuim.ie

Abstract: Perceptual centres (p-centres) are the subjective moments of occurrence of acoustic stimuli. When sounds are perceived in synchrony or are regularly spaced, it is their p-centres which occur synchronously or are isochronous. In order to analyse or model the acoustic features which influence the p-centre, it is necessary to measure p-centres for many stimuli. However there is a problem: it is difficult for an external observer to determine the exact time at which a listener perceives a sound's occurrence. A possible solution is to find a measurable electrophysiological correlate of the p-centre. In order to investigate this, an experiment is described which compares features of the Auditory Evoked Potential (AEP) response and p-centres for a number of speech and synthetic stimuli. The results indicate a correlation between the latency of the dominant negative peak of the AEP and the p-centre.

Introduction

The perceptual centre [1] or p-centre of a sound is defined as its perceptual moment of occurrence and is hypothesized to be a highly nonlinear function of the acoustic waveform. Researchers have for many years sought to measure p-centres for a wide variety of acoustic stimuli in an effort to derive robust models of the phenomenon. Unfortunately measuring p-centre data experimentally suffers from a fundamental methodological problem: how can an external observer determine the exact time at which a listener perceives a sound's occurrence?

A direct measure of the p-centre has been attempted by asking listeners to tap a finger in synchrony to the perceived beat of a stimulus presented in a repeating isochronous sequence, for example [2]. In order to synchronize a finger tap to a sound, the listener must anticipate the moment at which the sound will occur as the motor function takes some finite time to execute. Furthermore there is the question of what exactly the listener will synchronize. Sensory feedback for a finger tap may include vision, hearing and touch, each with its own sensory delay.

The principal method of indirectly measuring p-centres is via perceptual isochrony detection. A listener must determine whether or not a sequence of stimuli sounds perceptually isochronous. The most common approach is the method of adjustment applied to a pair of sounds A and B presented in a repeating duple

pattern with cyclic displacement, for example [3-5]. In this approach the cycle period is fixed and the listener must adjust the time of B to make the repeating sequence sound isochronous. The displacement of the onset time of the B from onset isochrony with A can be used to determine the p-centre of B relative to A. Hereinafter, this experimental approach will be referred to as Perceptual Isochrony Detection with Adjustment (PIDA). Unfortunately, subject bias, ability, human tolerance for limited anisochrony [6] and the subjective nature of the experiments have made clean data difficult to gather.

Some researchers have asked subjects to produce isochronous sequences of speech syllables, for example [7, 8]. In the absence of an isochronous reference, however it is difficult to measure p-centres from the resulting data. Deviation from onset isochrony can be used to calculate the p-centre of subsequent syllables relative to the first, but only if it is assumed that production is perfectly isochronous and that the isochronous interval is precisely known. If the subject is asked to produce speech synchronized to an isochronous reference, for example a metronome or click track, synchronization problems similar to those in tapping arise.

It would be desirable to gather large amounts of clean p-centre data for p-centre modelling, however it should be clear that there are difficulties with each of the methods described above. The novel approach undertaken here is to seek an electrophysiological correlate of the p-centre. This is achieved by measuring a listener's auditory evoked potentials (AEP) when presented with a variety of stimuli with different p-centre locations. Relative p-centres for the same stimuli are measured using the conventional method of perceptual isochrony detection. Data from the two measurement modalities can be compared in an effort to find features of the AEP that are correlated with relative p-centre measures. Such correlated features, if found, could be extremely significant for p-centre research, allowing quantitative measurements of the phenomenon in a more direct fashion than hitherto attempted. While there have been related studies investigating the electrophysiological correlates of speech onset perception [9] and rhythm perception [10], there does not appear to have been any investigation to date into the possible correlation between p-centres and the AEP.

Experiments were run with two different stimulus sets in an attempt to measure p-centre correlates in evoked potentials: the first with speech stimuli and the

second with synthetic stimuli. Thereafter, the results were analyzed to determine whether any features of the AEP co-varied with measured p-centres.

Materials and Methods

Two separate stimulus sets were used. The first of these consisted of four naturally spoken digits from 3 speakers (two male and one female) first described in [5] and summarised in Table 1.

Briefly, the speakers were asked to produce the digits at a speaking rate corresponding to a “marching pace” (approximately two words per second) and with a separation between words. This ensured that digit duration was sufficiently short to allow displacement without overlap in isochrony detection experiment. The stimuli were recorded as single channel audio with a sampling rate of 11025Hz and 16 bit resolution. The intensity of the samples was not equalized.

Table 1: Natural speech stimuli (stimulus set 1)

ID	Token	Duration (ms)	Speaker
OM1	One	490	Female
OM2	Two	355	
OM5	Five	442	
OM6	Six	470	
RV1	One	363	Male
RV2	Two	355	
RV5	Five	365	
RV6	Six	400	
TW1	One	585	Male
TW2	Two	405	
TW5	Five	575	
TW6	Six	490	

The second stimulus set consisted of six 1kHz tones, each amplitude modulated with a cosine shaped attack and release. The duration of each tone’s attack was different while the overall duration and release duration of each tone was fixed as summarized in Table 2.

The tones were synthesized as single channel audio with a sampling rate of 16000Hz and 16 bit resolution. The peak intensity of all tones was identical.

Table 2: Synthetic tone stimuli (stimulus set 2)

ID	Attack (ms)	Sustain (ms)	Release (ms)
A20	20	140	80
A40	40	120	80
A60	60	100	80
A80	80	80	80
A120	120	40	80
A160	160	0	80

Relative p-centres were measured in a quiet listening environment, seated at a desk equipped with a computer. Java software running on the computer, first

described in [5], managed the execution order of experimental trials, audio presentation within each trial and all subject interaction. User interface buttons allowed the subject start or stop audio presentation. The subject could adjust the sequence to make it sound more or less isochronous using the scroll wheel of a wheel mouse or buttons on the user interface.

The stimuli from stimulus set 1 were presented over loudspeakers at a comfortable level set by the subjects themselves. The stimuli from stimulus set 2 were presented monaurally (right ear) via an Eartone 3A insert earphone driven by the line out port of an M-Audio Duo D/A converter connected to the computer via USB. There was no analogue or digital volume control in the signal path. The presentation level was fixed by the digitized signal amplitude.

AEP measurements were made in a quiet listening environment that was not electrically shielded. Subjects were instrumented using a Biopac MP100 system. Recordings were between the vertex and the ear as is conventional for auditory evoked potentials. The amplifier gain was 50000 and the sampling rate was 2kHz.

For AEP measurement, all stimuli were presented monaurally (right ear) via the Eartone 3A insert earphone driven by the headphone port of an M-Audio Duo D/A converter connected to a notebook computer via USB. An analogue volume control on the headphone port allowed the subject to select a comfortable listening level.

For each stimulus a 2 channel audio signal was created. The right channel contained the stimulus audio and was directed to the insert earphone. The left channel contained a square wave pulse occurring just before stimulus onset. This pulse was amplified and routed to the Biopac MP100 to trigger acquisition of the AEP. This arrangement is depicted in Figure 1.

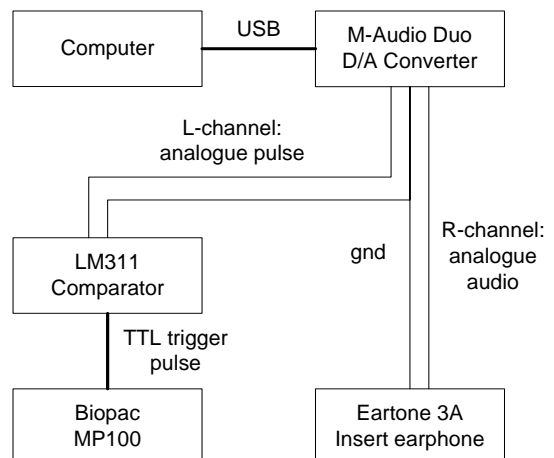


Figure 1: Triggering of AEP signal acquisition from 2 channel audio signal

The general procedure for measuring relative p-centres was PIDA, as described in [5]. Pairs of stimuli (A, B) were presented in a repeating duple pattern with

cyclic displacement of B. The subject task was to adjust the displacement of B to make the repeating sequence sound perceptually isochronous.

Stimuli in stimulus set 1 were grouped by speaker and the speaker's version of the "one" token was used as a reference. Each stimulus was paired with the appropriate reference in both A-B and B-A presentation orders. So for example OM2, OM5 and OM6 were each paired only with OM1. The nominal inter onset interval (IOI) for stimuli with identical p-centres was 600ms and deviation from the nominal IOI was used to calculate the relative p-centre.

In stimulus set 2, A20 was chosen as a reference stimulus. Each stimulus in the set (including A20 itself) was paired with the reference in both A-B and B-A presentation orders. The duration of stimuli in this set was shorter than those in stimulus set 1 and this allowed a shorter nominal IOI of 400ms without risking overlap of stimuli. In general a shorter IOI should improve the accuracy of p-centre measurements, as the just noticeable difference of isochrony is proportional to IOI [6].

AEP responses were recorded in the passive condition where the subject had no specific task to focus attention. The AEP measurements for each stimulus were based on 3 ensemble averages. Ensemble averages consisted of 500 trials each, so there were 1500 trials per stimulus. For each trial, AEP data collection was triggered at the start of stimulus presentation and 400ms data was recorded. Each trial incorporated a wait/rest period, so the duration of each trial was fixed at 1500ms. The trials in an ensemble followed one another without delay.

Two of the authors acted as subjects for AEP data collection. Three authors took part as subjects for relative p-centre measurements of stimulus set 1 while one author acted as the subject for relative p-centre measurements of stimulus set 2. For reasons of availability the latter subject was not one of the two authors who acted as subjects for AEP data collection.

Results

Grand average AEP data obtained for the stimuli in stimulus set 1 is shown in Figure 2, Figure 3 and Figure 4. Grand average AEP data for the ramped tones of stimulus set 2 is depicted in Figure 5. The dominant features of the AEP (the amplitude and time of its positive and negative going peaks) are straightforward to see. The most consistent feature appears to be the largest negative going peak which is located at times between approximately 150ms and 350ms after stimulus onset for the stimuli under test.

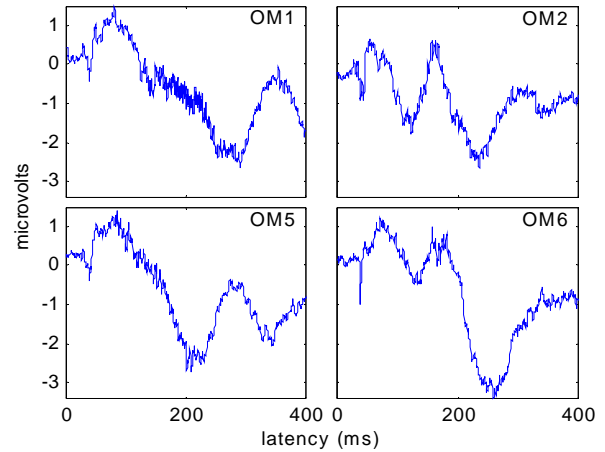


Figure 2: Grand average AEP data for stimuli from speaker OM

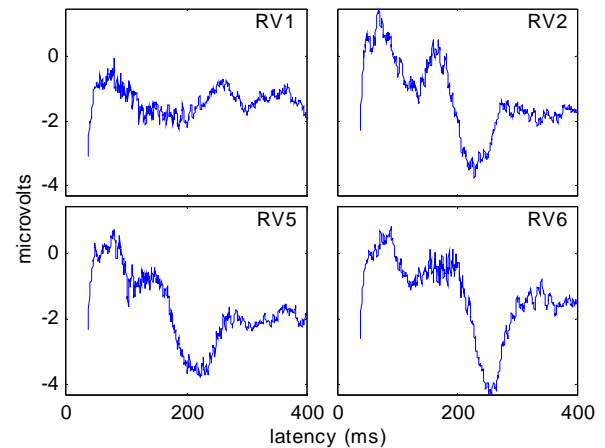


Figure 3: Grand average AEP data for stimuli from speaker RV. (Note that no AEP data was collected for the 38ms post stimulus onset for these stimuli.)

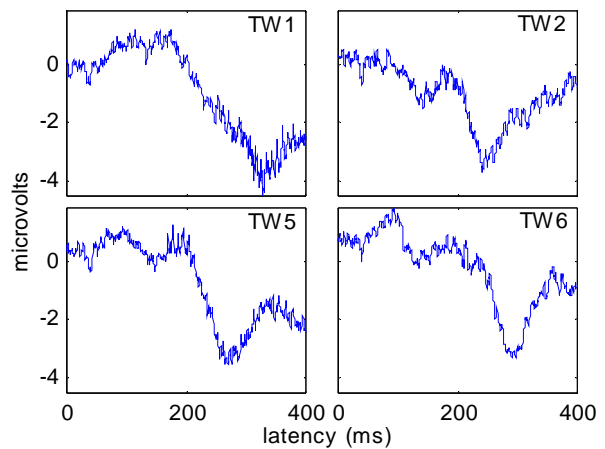


Figure 4: Grand average AEP data for stimuli from speaker TW

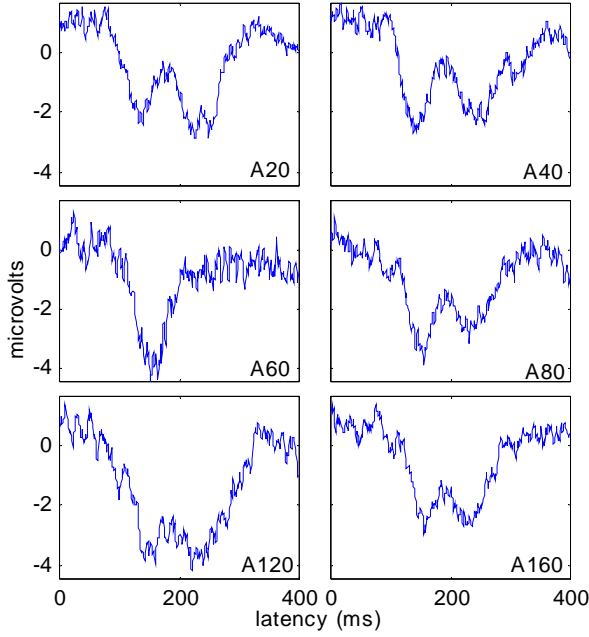


Figure 5: Grand average AEP data for stimulus set 2

Although ensemble averaging of AEP measurements emphasizes the signal and reduces the noise of a subject's AEP further processing is necessary to simplify extraction of the amplitude and time values for peaks and troughs. The smoothing method used here is a 15 point polynomial fit. The polynomial fit was found to give good smoothness without the potential phase shift or amplitude changes associated with low pass filtering as can be seen in Figure 6. The polynomial equation was differentiated to determine the location of the peaks and troughs for each stimulus.

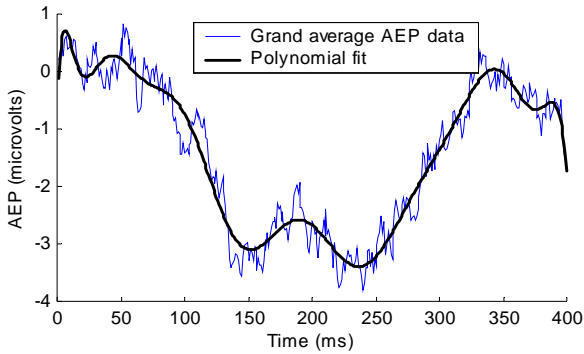


Figure 6: Grand average AEP data and corresponding polynomial fit for stimulus A120

The post onset latency of the deepest negative going peak, $t_B^{(N)}$, was noted for each stimulus. The AEP derived relative p-centre of stimulus B relative to stimulus A, P_{BA} , was then calculated according to equation 1.

$$P_{BA} = t_B^{(N)} - t_A^{(N)} \quad (1)$$

It should be obvious that if P_{BA} is positive, the latency of the absolute p-centre of B from the onset of B is longer than the latency of the p-centre of A from the onset of A.

Each PIDA trial for a pair of stimuli A and B recorded the displacement, d_{BA} , of the onset of B from the time of onset isochrony with A required to make the repeating sequence sound perceptually isochronous. P_{BA} and d_{BA} are trivially related by equation 2.

$$P_{BA} = -d_{BA} \quad (2)$$

Using equations 1 and 2 equivalent relative p-centres can be derived from AEP data and from perceptual isochrony detection data. These calculated values are listed in Table 3 and Table 4.

Table 3: Comparison of measured p-centres and those derived from the dominant negative peak in AEP for stimulus set 1

Relative P-centre	PIDA derived (ms)	AEP derived (ms)
$P_{OM2-OM1}$	-65	-41
$P_{OM5-OM1}$	-76	-57
$P_{OM6-OM1}$	-12	-33
$P_{RV2-RV1}$	27	37
$P_{RV5-RV1}$	9	24
$P_{RV6-RV1}$	57	64
$P_{TW2-TW1}$	-187	-83
$P_{TW5-TW1}$	-91	-59
$P_{TW6-TW1}$	-113	-31

Table 4: Comparison of measured p-centres and those derived from the dominant negative peak in AEP for stimulus set 2

Relative P-centre	PIDA derived (ms)	AEP derived (ms)
$P_{A40-A20}$	-2	8
$P_{A60-A20}$	6	20
$P_{A80-A20}$	14	17
$P_{A120-A20}$	16	18
$P_{A160-A20}$	32	25

The relative p-centres of the speech stimuli from stimulus set 1 derived from PIDA measurements are correlated with those derived from AEP measurements ($r = 0.9$) and this correlation is highly significant ($p < 0.001$). The principal outliers in the data relate to speaker TW whose tokens were substantially longer than other speakers. Subjects participating in the PIDA trials reported difficulty in achieving perceptual isochrony with these stimuli. It should also be noted the

measured relative p-centres above were collected with an inter onset interval of 600ms and that the just noticeable difference for isochrony detection with this inter-onset interval is ± 15 ms [6].

The relative p-centres for the ramped tones in stimulus set 2 derived from PIDA measurements are correlated with those derived from AEP measurements ($r = 0.85$), but the correlation is not statistically significant ($p = 0.066$). The standard deviation of individual PIDA trial results ranged from 10ms to 19ms and was highly correlated with the attack time of the tones ($r = 0.92$). The correlation was significant ($p < 0.01$).

Discussion

There are differences between the two stimulus sets which may have an effect on the AEPs. The speech stimuli are naturally wideband signals while the tones are clearly narrowband signals. The AEP amplitude is consistently larger for the tone stimuli than for the speech stimuli.

P-centres measured using perceptual isochrony detection are usually averaged across multiple subjects on the assumption that individual subjects do not vary significantly in their perception of the p-centre or isochrony. In contrast, the AEP data presented in this paper was recorded from either one or two subjects. It is possible that additional insights might be gathered by comparing the PIDA data and AEP data for a single subject. In particular it is worth remembering that the AEP data and PIDA data for stimulus set 2 were recorded from different single subjects, so it is possible that the non-significance of the correlation between the two data sets was due to subjective differences.

Nevertheless the results presented above are suggestive of a relationship between the dominant negative peak of the AEP and the p-centre of a sound in the case of monosyllabic speech. On the other hand, no conclusion can be drawn on the relationship in the case of ramped tones.

References

- [1] J. Morton, S. Marcus, and C. Frankish, "Perceptual Centers (P-centers)," *Psychological Review*, vol. 83, pp. 405-408, 1976.
- [2] P. M. Janker, "The range of subjective simultaneousness in tapping experiments with speech stimuli," presented at ESCA Workshop on the Auditory Basis of Speech Perception, Keele University (UK), 1996.
- [3] S. M. Marcus, "Acoustic determinants of Perceptual-center (P-Center) location," *Perception and Psychophysics*, vol. 30, pp. 247-256, 1981.
- [4] S. K. Scott, "The point of P-centres," *Psychological Research*, vol. 61, pp. 4-11, 1998.
- [5] R. Villing, T. Ward, and J. Timoney, "P-Centre Extraction from Speech: the need for a more reliable

measure," presented at Irish Signals and Systems Conference, Limerick, 2003.

- [6] A. Friberg and J. Sundberg, "Time discrimination in a monotonic, isochronous sequence," *Journal of the Acoustical Society of America*, vol. 98, pp. 2524-2531, 1995.
- [7] A. D. Patel, A. Lofqvist, and W. Naito, "The acoustics and kinematics of regularly timed speech: a database and method for the study of the P-Centre problem," presented at 14th International Congress of Phonetic Sciences, San Francisco, 1999.
- [8] C. A. Fowler, "'Perceptual centers' in speech production and perception," *Perception and Psychophysics*, vol. 25, pp. 375-388, 1979.
- [9] L. D. Sanders, E. L. Newport, and H. J. Neville, "Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech," *Nature Neuroscience*, vol. 5, pp. 700-703, 2002.
- [10] P. Fraisse and A. Lavit, "Auditory evoked potentials and the perception of rhythms," *International Journal of Psychophysiology*, vol. 4, pp. 85-89, 1986.